

AD A091621

LEVEL 12

Technical Note

1980-39

Design and Development
of a Seismic Data Center

M. A. Chinnery
A. G. Gann

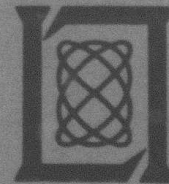
11 August 1980

Prepared for the Defense Advanced Research Projects Agency
under Electronic Systems Division Contract F19628-80-C-0002 by

Lincoln Laboratory

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LEXINGTON, MASSACHUSETTS



Approved for public release; distribution unlimited.

DDC FILE COPY

80 11 17 096

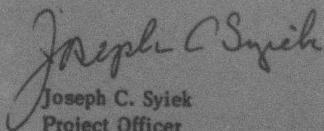
The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology. This work was sponsored by the Defense Advanced Research Projects Agency under Air Force Contract D19628-80-C-0002 (ARPA Order 512).

This report may be reproduced to satisfy needs of U.S. Government agencies.

The views and conclusions contained in this document are those of the contractor and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the United States Government.

This technical report has been reviewed and is approved for publication.

FOR THE COMMANDER



Joseph C. Syiek
Project Officer
Lincoln Laboratory Project Office

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
LINCOLN LABORATORY

DESIGN AND DEVELOPMENT OF A SEISMIC DATA CENTER

M. A. CHINNERY
A. G. GANN
Group 22



TECHNICAL NOTE 1980-39

11 AUGUST 1980

Approved for public release; distribution unlimited.

LEXINGTON

MASSACHUSETTS

ABSTRACT

The U.S. has embarked on the design and development of a sophisticated data center for the analysis and management of seismic data. There are two broad motivations for this project. First, we need to develop a new generation system for the management and retrieval of digital seismic data, and to provide a modern data resource for the research community. Second, we need a facility that will be able to respond to any U.S. obligations that might be incurred under future agreements on international seismic data exchange and global seismic monitoring, such as those currently under discussion by the Committee on Disarmament. A computer architecture has been selected, consisting of a series of minicomputers connected by a high data rate local computer network. The principal components of the system are discussed, with emphasis on the data base management system, and the seismic analysis station, which provides an interactive man-machine interface.

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Special

TABLE OF CONTENTS

I.	INTRODUCTION	1
II.	OBJECTIVES	1
III.	APPROACH	2
IV.	FUNCTIONAL REQUIREMENTS FOR THE DATA CENTER DESIGN	3
	A. Input Data Characteristics	6
	B. Data Storage and Retrieval Requirements	8
	C. Routine Output Products	9
	D. The User Interface	11
	E. Internal Processing Requirements	12
V.	DATA CENTER OPTIONS	17
VI.	SYSTEM ARCHITECTURE	18
	A. Chosen Architecture	18
	B. Subsystems	21
	C. Computer Hardware	23
VII.	SUBSYSTEM DESCRIPTIONS	24
	A. Seismic Analysis Station	24
	B. Database Subsystem	27
	C. Research Support Subsystem	36
	D. System Control and Services Subsystem	37
	E. Automatic Processing Subsystem	37
	F. Communications Interface Subsystem	38
	G. Local Computer Network	39

I. INTRODUCTION

The U.S. has embarked on the design and development of a sophisticated data center for the analysis and management of seismic data. There are two broad motivations for this project. First, we need to develop a new generation of seismic data management and analysis system that will utilize digital seismic data and provide a modern data resource for the research community. Second, we need a facility that will be able to respond to any U.S. obligations that might be incurred under future agreements on international seismic data exchange and global seismic monitoring, such as those currently under discussion by the Committee on Disarmament.

In this report we formulate the basic objectives of this Seismic Data Center (SDC), outline the functions that we anticipate it will be required to perform, and describe a computer architecture that has been selected as a basis for further development.

II. OBJECTIVES

Objectives of the data center fall into two classes, corresponding to the two broad motivations mentioned above. The objectives in support of the seismic research community are as follows:

1. To provide a state-of-the-art seismic data management facility for the storage and retrieval of digital seismic data.

2. To provide a test-bed for the evaluation of a variety of automatic signal processing techniques for digital seismic data.
3. To develop the most efficient techniques for the rapid preparation of a comprehensive event bulletin, using both digital waveform data and station seismic arrival reports.
4. To provide a variety of services to the research community, including both local and remote access to all archived data and event bulletins.

The objectives in support of international exchange of seismic data are as follows:

1. To provide a facility that will meet all anticipated requirements for international exchange of seismic data.
2. To provide a test bed for evaluating the hardware, software and processing techniques that may be required under contemplated agreements on international exchange of seismic data.

III. APPROACH

It is not possible, at present, to formulate a complete set of the functions that this data center will be required to perform. New digital stations are still being installed, and considerable research is needed before we can fully understand the best ways to use this new type of data. In the international arena, agreements on global seismic monitoring have

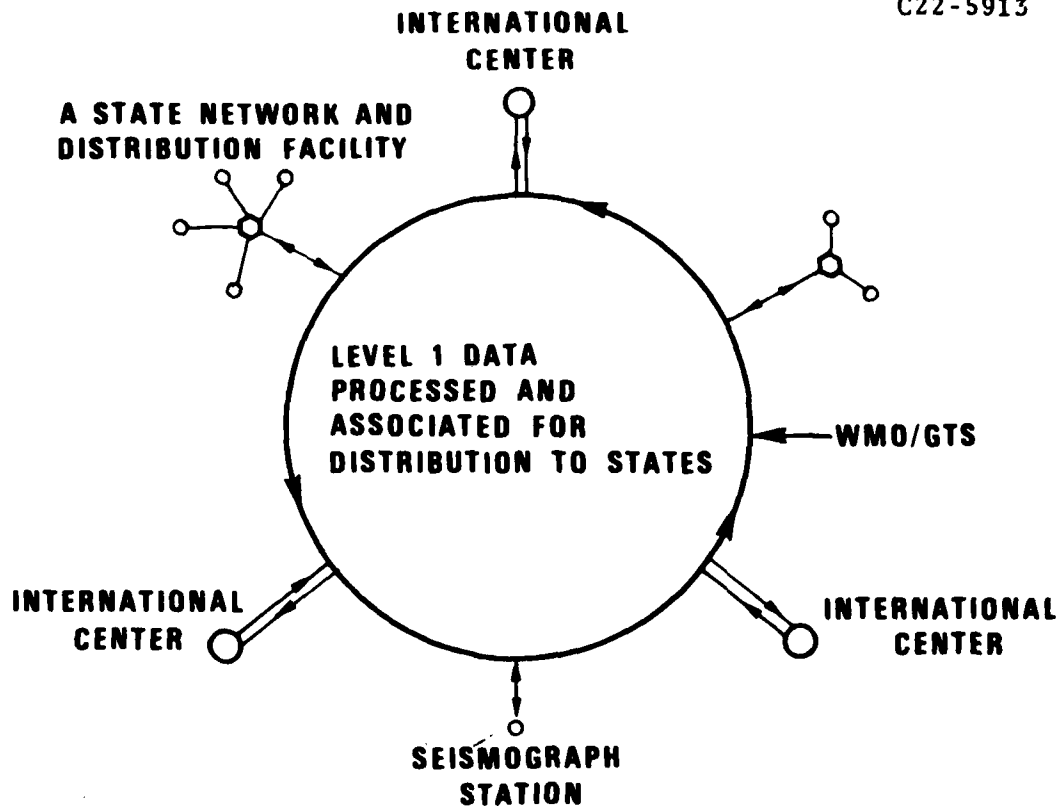
yet to be concluded, and we cannot predict the detailed requirements that these agreements will place on the center.

We have, therefore, initiated a development plan aimed at testing the feasibility of data management and analysis concepts. Initial work is aimed at the formulation of a computer hardware and software architecture which will be flexible enough to provide a test-bed for analysis procedures, and will enable us to study the optimum application of modern technology to the data center problem.

This report describes our first prototype system. As it is being developed, it has a limited data capacity, but is readily expansible to a full scale system when needed. We have formulated a preliminary set of functional requirements, on which this prototype system is based, recognizing that some of these requirements are likely to change as time passes.

IV. FUNCTIONAL REQUIREMENTS FOR THE DATA CENTER DESIGN

The Seismic Data Center will have to carry out two broad classes of functions, corresponding to the requirements of the research community, and the requirements of the international community. Because of the large quantity of information involved, the design of the data center will be dominated by the research requirement for the storage and retrieval of digital waveform data, and the processing of these data into a comprehensive event bulletin. It is convenient to consider the international requirements as a sub-set of this overall design task; however, we



PURPOSE OF STATE FACILITY:

- 1) SENDS LEVEL 1 DATA
- 2) RECEIVES EPICENTERS AND ASSOCIATED IDENTIFICATION PARAMETERS
- 3) SENDS AND RECEIVES LEVEL 2 DATA ON REQUEST

A SINGLE STATION WHEN OFFICIALLY AUTHORIZED HAS SIMILAR COMMITMENTS

Fig. 1. Data flow in an international exchange of seismic data.

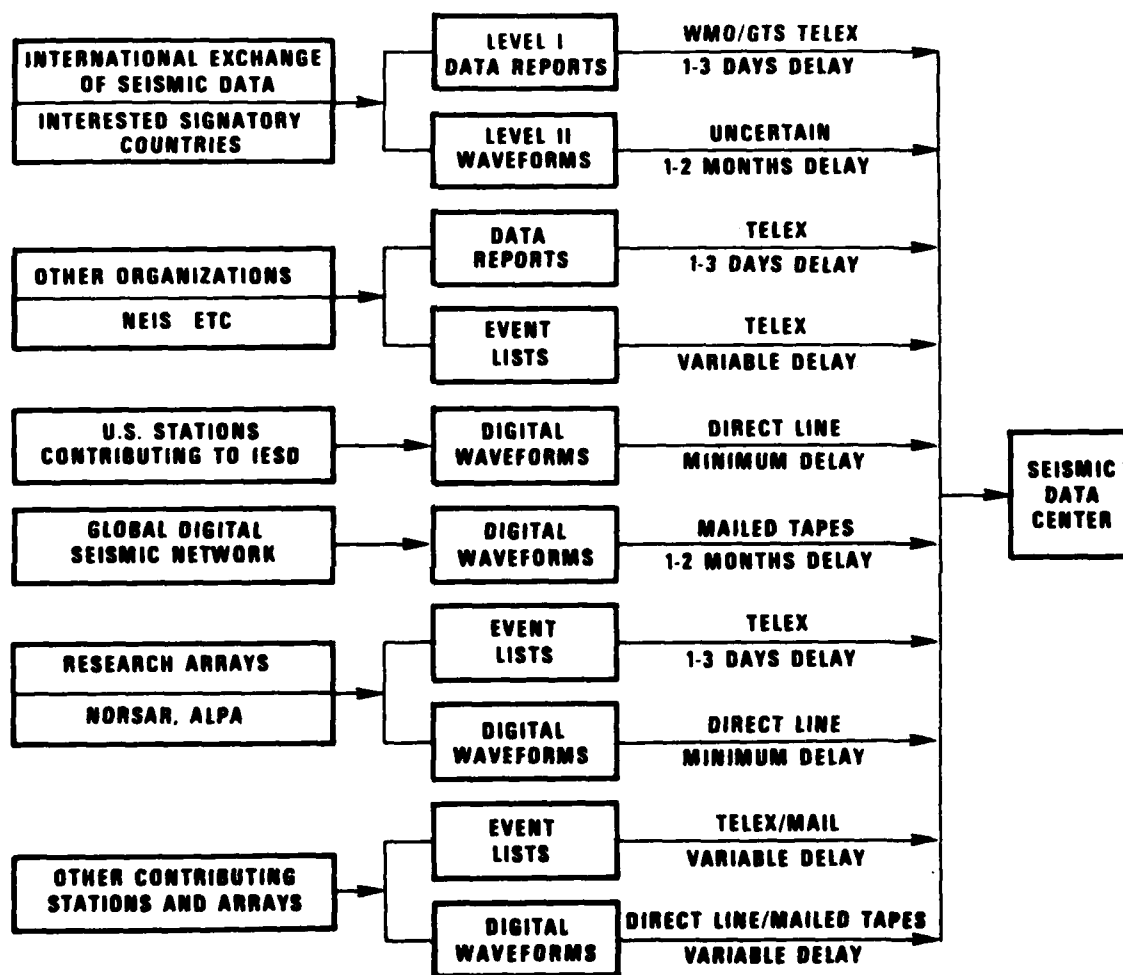


Fig. 2. Seismic data center input data streams.

recognize that they may differ so much from the research requirements that it may be necessary to carry them out in a separate International Data Center Facility.

The requirements of the seismic research community are formulated in a preliminary way below. Discussions with the seismic community are continuing in an effort to refine these requirements. Requirements for International Exchange of Seismic Data (IESD) are taken from reports of the CD Ad Hoc Group of Experts, particularly documents CCD/558 (9 March 1978) and CD/43 (25 July 1979). The proposed scheme for IESD is shown in Figure 1 (taken from CCD/558).

The total range of SDC functional requirements is discussed in the following subsections.

IV.A. Input Data Characteristics

The Seismic Data Center will be required to accept data from a wide variety of sources. Figure 2 summarizes these sources.

Three types of data are anticipated. The first is data reports (the level I data described in CCD/558 and CD/43). These are measurements made by station operators on observed seismic arrivals, and include a variety of parameters, arrival time, amplitude and phase identification (a comprehensive list of these parameters has been given in CD/43). The second is digital waveform data from both single stations and arrays. Some of these

stations may be U.S. contributions to IESD; for these stations, one task of the data center will be to extract data reports from the waveforms, and distribute them to all IESD participants. The third may consist of actual lists of events, with preliminary event parameters. Such lists are frequently prepared by array stations and certain existing earthquake information organizations.

These data types will arrive at the SDC in a variety of different ways, and with highly variable delays (see Figure 2). One significant task of the SDC will be to ensure that each of these sets of data is properly entered into the SDC data base.

The rates at which these data enter the system will place an important constraint on the system design. A modern three component digital seismic station may generate data at a rate in excess of 2 kbits/second. If a network of such stations were to feed continuous data into the Seismic Data Center, this data source would dominate all others by far, in terms of both processing and storage requirements. We have selected a system design input rate of 125 kbits/second, since this would appear to provide ample capacity for the foreseeable future. We also require that the system be able to handle twice this data rate during catch-up periods (e.g., following a hardware failure, and that the system could be adapted to an average input rate of 250 kbits/second by simply adding more hardware, without invalidating the existing hardware.

IV.B. Data Storage and Retrieval Requirements

Requirements on the data base management system concern its capacity and its speed of retrieval. The design data rate of 125 kbits/second corresponds to 10^{10} bits/day. This requirement can be met by a combination of on-line disk storage and off-line tape storage (commercially available disk units can store about 5×10^9 bits, and high density tape units can store about 10^9 bits on a 2400 foot tape). We have selected this tape/disk solution for early SDC development. We also recognize that other systems (such as optical disks) may become available within the next few years. The SDC architecture must have the flexibility that such devices can be incorporated into the system with little problem at a later stage.

The required rate of retrieval of data from on-line disk storage depends somewhat on the type of processing carried out at the SDC. We do not anticipate that there will normally be a requirement to retrieve large quantities (say, all data for an entire day) in one transmission. However, in order not to exclude this possibility, we require that data can be moved from disk storage to other parts of the SDC at a minimum of an order of magnitude faster than real time (this corresponds to a net signalling rate of at least 2 Mbits/second). With this signalling rate, small quantities of data (say, a few waveforms), can be retrieved from disk storage in a fraction of a second.

Retrieval of data from off-line tape storage will inevitably take

longer. We, therefore, set the requirement that, under normal conditions, all input data will be stored on-line until analysis of that data is completed. Retrieval from tape will then be limited to a small number of requests, primarily from research scientists, and we require simply that all the waveforms for a given event be on a single tape. The time needed to respond to a request for such data will then be the time necessary to select a tape, mount it and read it into the system. Depending on the total volume of data input to the center, each tape will contain all available data for a period ranging from 2 hours to 1 day.

In addition to routine functions, we also require that the data base have the capacity to store certain special data sets on-line. These would include reference events, used to aid an analyst in seismic interpretation, and data sets accumulated in response to a user request.

IV.C. Routine Output Products

The SDC will produce a wide variety of routine output data and provide a range of user services. These are summarized in Figure 3.

There are two principal routine products. First, the SDC must produce a comprehensive event bulletin. This will provide a data base in itself for some users; to others, it will function as an index to the waveform data stored at the SDC. Document CCD/558 suggests that this bulletin should be available with a net delay of 3-5 days from real time. A bulletin produced this quickly will also be of substantial use to the research

community. However, we note that some data may not be available at the SDC on this time scale, and this suggests that it may be necessary to prepare several bulletins. We require that the SDC be able to compile an event bulletin within 24 hours of the receipt of the last piece of input data used in its preparation (this assumes 7 days a week SDC operation). This constraint will satisfy all anticipated requirements.

The contents of the event bulletin are less well defined. Certainly, the bulletin must contain all the event parameters currently included in such publications as the Earthquake Data Reports (of the U.S. Geological Survey). We anticipate that current and future research may lead to substantial improvements in this list of parameters, and parameters relating, for example, to source mechanism may be included. One immediate addition will be information about the availability of waveform data for listed seismic arrivals.

The second important routine product will be access to the waveform data base. We anticipate that user requests for waveform data will be of two kinds. The user may request waveform data for a particular event, as listed in the event bulletin; or he may request the waveform data from a given station (or stations) for a given time interval. Where building up a large data set, he may leave a standing order with the SDC for certain specific event types.

IESD output products will follow specifications laid out in interna-

tional agreements. At present, we anticipate general distribution of data reports from U.S. stations which are included in IESD, and an event bulletin, both via the WMO/GTS network. Level II waveform data will be supplied by whatever methods are developed by the CD.

IV.D. The User Interface

We plan to provide a variety of methods whereby the user can interact with the data center. For those who can visit the center, in-house computer facilities will be provided for interaction with the data base, local interactive display and analysis.

For users who are unable to reach the center, several systems will be provided. Data requests sent by mail or telephone will be satisfied by mailed tapes or transmission over a low data rate channel, such as a dial-up phone connection. The latter will be adequate for the transmission of event bulletins, or small quantities of digital data.

Where larger amounts of data are required, a more sophisticated form of access would be needed for local recording and computational facilities. We anticipate that the capability required would range from a simple micro-computer based terminal to something like a Seismic Analysis Station, described later in this report, depending on the level of support required. Such a terminal would be connected to the data center by a high data rate communications link.

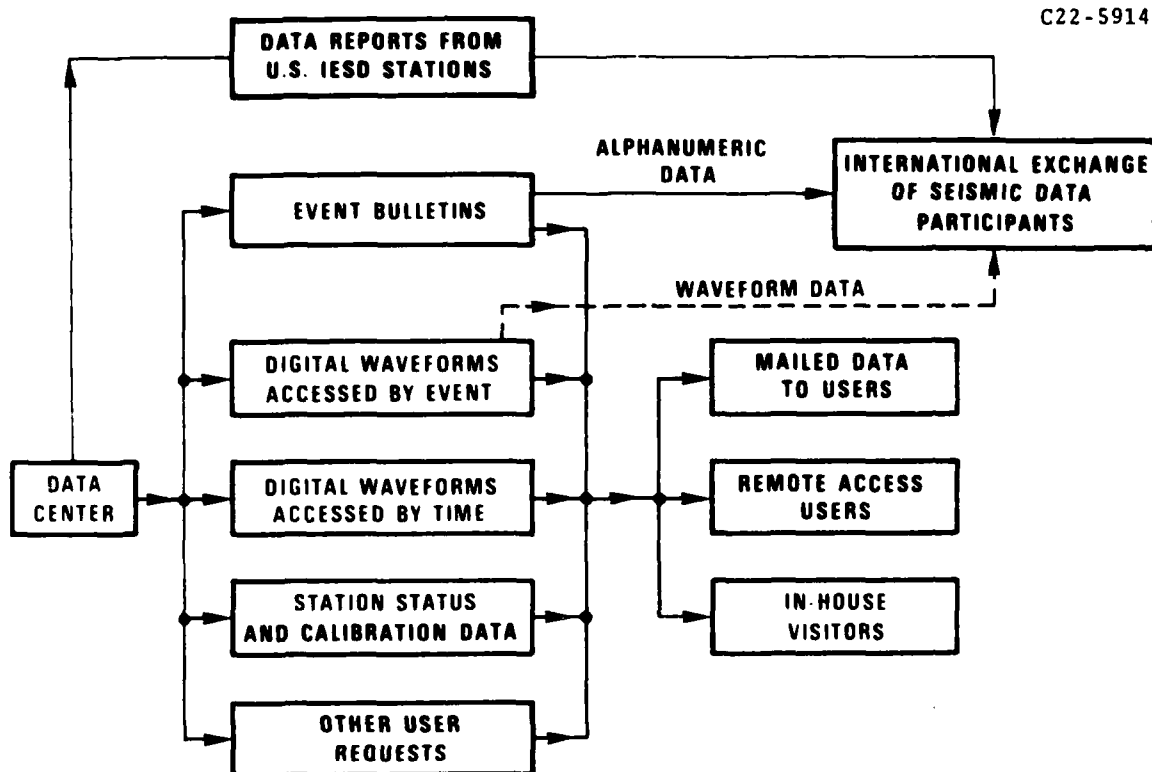
Certain types of analytical tools will also be made available to the user. Examples include searching the event bulletin on one or more of the event parameters, carrying out filtering, spectral analysis, rotation of horizontal components and other kinds of time series analysis, and analytically modifying the instrument response of the stored data (e.g., to produce a broad-band signal). Within the limitations of system capacity, users will also be able to apply user-generated processing methods to the data.

IV.E. Internal Processing Requirements

In order to translate the input data (Figure 2) into the output data products (Figure 3), it will be necessary to carry out a series of seismic processing steps within the SDC. These processing steps will set some important requirements on the SDC design.

Figure 4 shows, in schematic form, the processing steps involved in the preparation of a final event bulletin. There are two basic sets of procedures. The first is carried out completely automatically on the input data streams; the second is carried out by a human analyst, who will refine the results of the automatic processing steps using a variety of software tools and with access to all available data.

The actual processing tasks that will be carried out automatically are not yet clear. Many existing systems include only signal detection in this category; however, there are many indications in the literature that parameters such as arrival time, amplitude, dominant period and signal character



Figs. 3 Data center products and user interfaces.

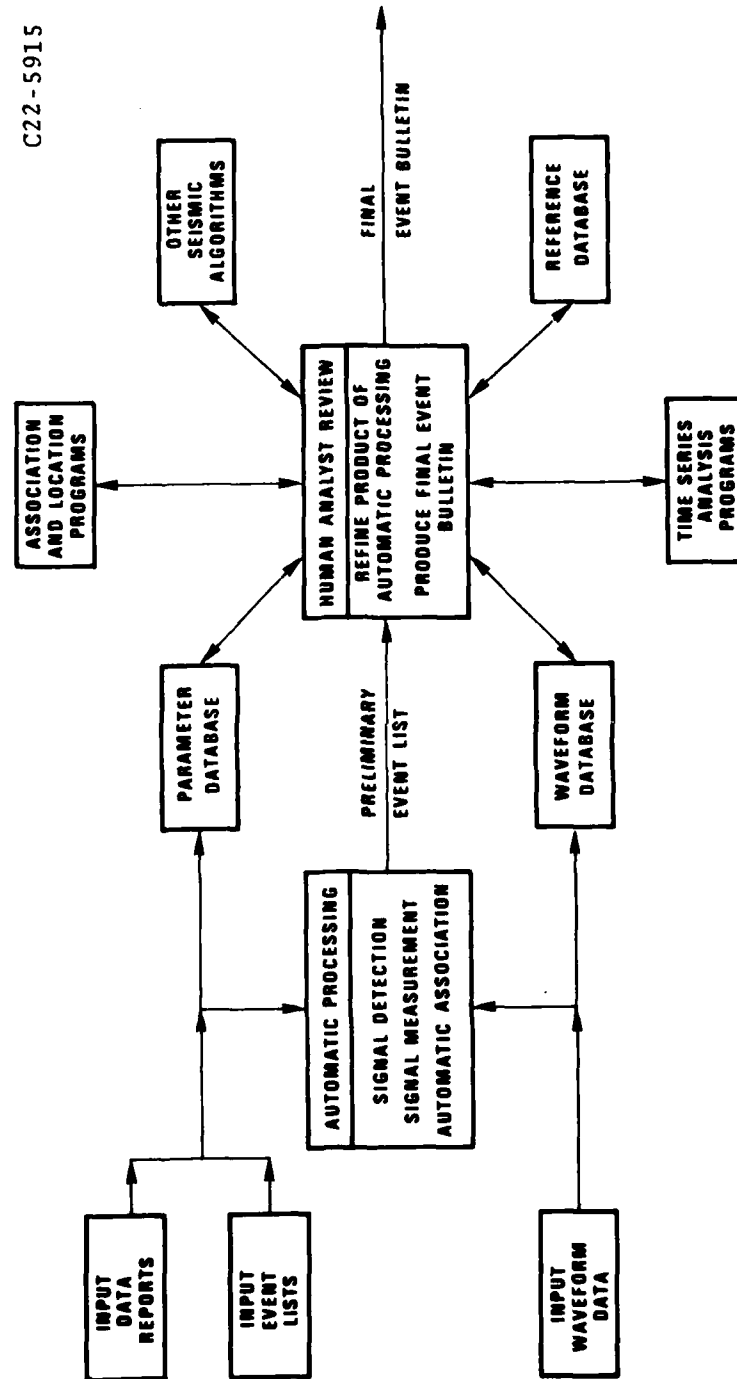


Fig. 4. Data center functions and procedures.

can usefully be measured by the application of automatic algorithms. A program of research is being carried out in the U.S. with the objective of evaluating and implementing these concepts. We anticipate that this will substantially reduce the amount of human analyst time necessary in event bulletin preparation.

Based on our present experience, however, there is no way in which the human analyst can be eliminated from the processing procedures. The subjective judgement and experience of the mind are essential to the production of an acceptable event bulletin. This places a requirement on the system for a sophisticated interactive man-machine interface. The analyst must be able to interact directly with event lists, lists of signal characterization parameters, and digital waveforms. He must be supplied with a wide variety of software tools, including time series analysis programs, association and location programs, and programs to aid him in the determination of event parameters such as magnitude and source mechanism.

Such an interactive "seismic analyst station" will include all of the basic features that a researcher will need in order to interact with the system. Our development of a seismic analyst station is, therefore, aimed at devising a very general purpose interactive seismic terminal, which (with minor software variations) will satisfy the needs of the operational analyst, the visiting research scientist and the remote terminal operator.

<u>DATA CENTER TYPE</u>	<u>FUNCTIONS</u>	<u>REQUIRED FACILITIES</u>
INTERNATIONAL CENTER (MINIMAL)	RECEIVE LEVEL I DATA TRANSMIT/RECEIVE LEVEL II DATA PREPARES EVENT BULLETINS	TELETYPE WMO/GTS CONNECTION TELECOPIER
INTERNATIONAL CENTER	AS ABOVE WITH LIMITED DIGITAL DATA	MINICOMPUTER, MODEST PERIPHERALS WMO/GTS CONNECTION TELECOPIER
INTERNATIONAL CENTER AND STATE FACILITY	RECEIVE, PROCESS AND STORE DATA FROM STATE NETWORK ALL INTERNATIONAL CENTER FUNCTIONS	DATA COMMUNICATIONS INTERFACE AT LEAST 2 MINICOMPUTERS WITH DISK/TAPE STORAGE TELETYPE, TELECOPIER
REMOTE RESEARCH SITE (MINIMAL)	ACCESS SDC DATA BASE	SIMPLE TERMINAL PHONE LINE CONNECTION
REMOTE RESEARCH SITE	INTERACT WITH SDC DATA BASE	INTERACTIVE COMPUTER SYSTEM HIGH DATA RATE LINE
FULL SCALE SEISMIC DATA CENTER	RECEIVE, PROCESS AND STORE ALL AVAILABLE DATA ALL INTERNATIONAL CENTER FUNCTIONS EXTENSIVE USER SERVICES	AS DESCRIBED IN THIS REPORT

Fig. 5. Some data center options for IESD participation.

V. DATA CENTER OPTIONS

Although the Seismic Data Center described in this report is a very sophisticated and expensive high capacity system, we recognize that many research and international organizations may have more limited overall requirements, and may desire a more modest facility. Figure 5 lays out a series of system options that should be considered by a potential user.

The basic requirements of a CD International Center (Figure 1) are very limited. Level I data (alphanumeric data reports) and event bulletins may be received with a standard inexpensive teletype machine connected to the WMO/GTS network. Receipt and transmittal of small amounts of level II (waveform) data may be accomplished using a telecopier or similar device. A more modern facility, still quite inexpensive, could be developed by the addition of a minicomputer with limited peripherals, including tapes, disks and standard terminals.

The research user might have a variety of needs, ranging from a simple terminal dial-up connection for the receipt of event bulletins and the ordering of mailed tape data, to a sophisticated interactive terminal such as the seismic analysis station described below.

Complications, and expense, only begin to arise when the system is required to receive, process and store a substantial amount of digital waveform data, which might happen at a State Network and Distribution Facility (Figure 1). The technology for this system has been well

developed in many countries as part of the earthquake prediction program. Typical installations include at least two mini-computers (one for data receipt and tape storage, and one for processing) together with a variety of peripheral devices.

The Data Center described here is the highest capacity system, designed not only to carry out all of the above tasks, but to provide sophisticated processing, rapid data retrieval and a variety of user services. We anticipate that the techniques developed in the full-scale SDC will have applications to each of the more modest systems.

VI. SYSTEM ARCHITECTURE

In this chapter we describe, in very general terms, the overall architecture of the system. In the first section, the chosen architecture is discussed, and the major subsystems defined. Subsequent sections give an overview of the functions of each subsystem. The next chapter will give more detail on the design of the subsystems.

VI.A. Chosen Architecture

There are many options from which to choose the computer architecture of the Seismic Data Center. The first choice is between:

1. A single mainframe computer, perhaps consisting of a central computer and multiple peripheral processors, and

2. Some form of distributed processing system in which the facility consists of a community of interconnected computers.

If the single mainframe option were selected then on the basis of reliability the single mainframe would need to be duplicated to prevent a single failure from disabling the entire system. This would double the cost with no corresponding increase in useful performance. There are also difficulties with a single mainframe on the basis of flexibility. A single computer system must be sized for the largest data and computation load. If that load does not materialize the cost will have been considerably more than for a properly sized system.

However, if the load exceeds the capability of the single mainframe, the operational and system design perturbations can be major. The need for high reliability and for a flexible system which could be adapted to a wide range of requirements has lead to the selection of a multiple interconnected computer architecture for the SDC. The individual units can be of modest size and the number of units can be adapted to satisfy operational requirements.

Given that a number of interconnected computers are to be used, it still remains to decide if they are to be tightly integrated to constitute a single virtual machine which will support the SDC or if they will be more loosely coupled into some form of computer network with different SDC tasks identified with the different machines. It is the second option which has

C22-5917

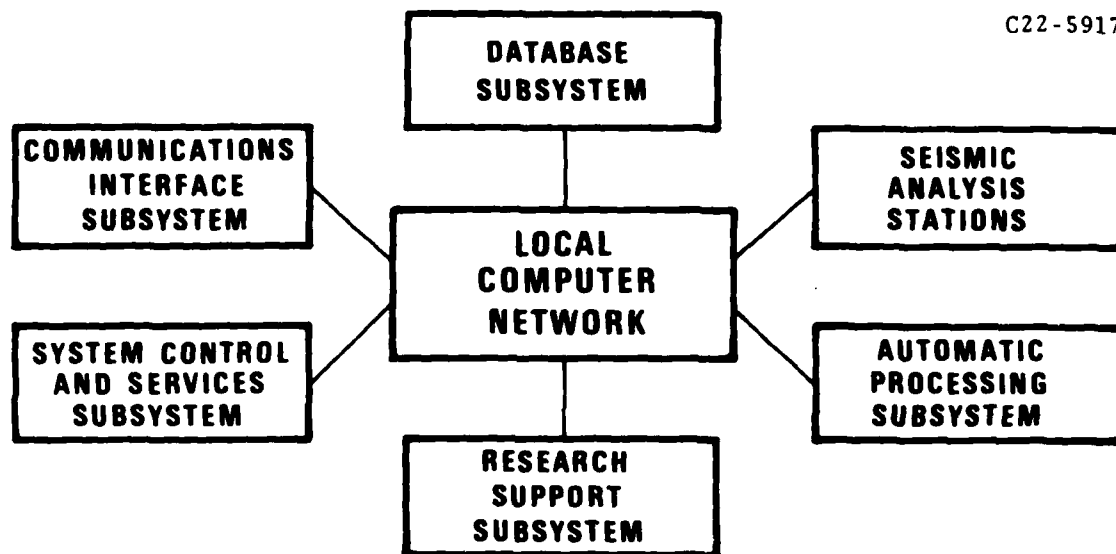


Fig. 6. Conceptual seismic data center configuration.

been selected for the SDC design. The first one implies the need for more specialized software and distributed processing capabilities. Such systems are necessarily very complex and are in the domain of advanced research rather than system development.

The second requires a local computer network which can allow existing systems to communicate data and command/status information without forcing a tight integration of the operating systems. The second also gives more overall flexibility and easy expansibility without the need for extensive software reorganization.

VI.B. Subsystems

Figure 6 shows the loosely coupled architecture which has been chosen. In this figure are shown the major components of the computer system which forms the central element in the Seismic Data Center. These components are subsystems generally consisting of one or more computers with attached peripherals arranged around the local computer network subsystem, which is the key to the effectiveness of the chosen architecture. The other subsystems are:

- * The Seismic Analysis Stations which provide the seismic analysts and users with the ability to perform extensive computations on both the waveforms and the elements of the parameter database.

- * The Database Subsystem which consists of the Parameter Database and the Waveform Database. The Parameter Database stores all of the alphanumeric data and provides a data management and retrieval system for events and arrivals. The Waveform Database provides a facility for maintaining and accessing the very large amounts of waveform data which the system acquires and processes.
- * The Research Support Subsystem which will carry out a variety of computational tasks in support of research users, both local and remote. This support will include the selection of data sets for study, the analysis of both parametric and waveform data and the development of new algorithms and computer programs.
- * The System Control and Services Subsystem which will carry out a variety of tasks. These tasks include the preparation of data for distribution as requested by authorized outside users, development of new software for the system, and providing other basic system services.
- * The Automatic Processing Subsystem which will carry out automatic (non-interactive) computational tasks related to detecting the arrival of seismic signals from the digital waveform data and the automatic association of arrival information into the preliminary characterization of events.

- * The Communications Interface Subsystem which is responsible for all of data input, and certain types of data output.
- * A Local Computer Network Subsystem which interconnects the computers and provides all intercomputer communication of data and control information. The local computer network is a key element in the system. It is required to be extremely reliable since it carries all intercomputer communications including all the data and all the command and status traffic in the system.

VI.C. Computer Hardware

The minicomputer chosen to implement the architecture is the Digital Equipment Corporation's PDP-11 (tm Digital Equipment Corp.) family. One of the principal factors in the choice of the PDP-11 family for the basis of the SDC computer system is the existence of the UNIX (tm Bell Labs) operating system for the various members of this family of computers. The UNIX operating system has been successfully used by the Lincoln Laboratory Applied Seismology Group to support the group's on-going seismological research program. UNIX has been adopted for use by a number of other seismological research groups. One result is that a number of seismic data handling and processing application programs have been developed for the UNIX system.

In addition, the UNIX user interface provides a good environment for interactive data manipulation and analysis, and for the development of

libraries of applications programs which can be applied to data analysis in a very flexible way. Since the cost of software has become the overwhelmingly predominant item in the cost of most modern systems employing computers, the decision to use UNIX as the basis for the user interface was made very early in the design process.

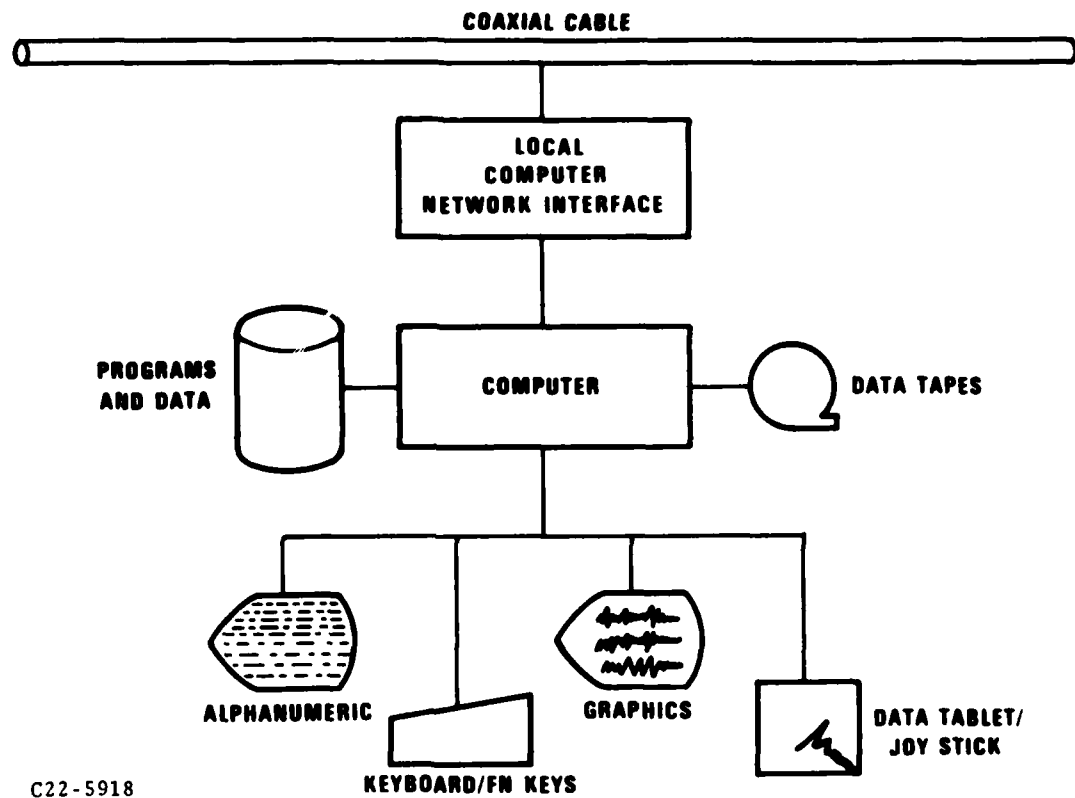
VII. SUBSYSTEM DESCRIPTIONS

The following sections give individual descriptions of the subsystems introduced in the previous chapter.

VII.A. Seismic Analysis Station

The Seismic Analysis Stations provide the normal user interface to SDC computer system. The Seismic Analysis Stations are used by the analysts to do all processing to prepare the event bulletin. They are also used to do all interactive waveform processing and to support research. The following is an overview of the design.

The Seismic Analysis Station (SAS) is a dedicated interactive computer system based on a PDP-11 series computer running the UNIX operating system. A representation of the configuration of the SAS is given in Figure 7. The user interface provides a convenient way for the analyst to call up, display, and manipulate waveform data and parameter data, and to run a variety of processing routines on both the waveform and the parametric data.



C22-5918

Fig. 7. Seismic analysis station configuration.

The user communication with the system will be through two types of displays, one for waveform data and one for alphanumeric data, along with a keyboard and a data tablet and/or other analog input devices. Each SAS will be connected to the Local Computer Network through an interface to allow rapid access to the two database subsystems. The SAS will also have a significant size local disk storage capability on which the analyst will keep the data to which he is actively referring and manipulating.

The normal operating procedure will be to specify a segment of time with which the operator is going to deal and the database will then transfer the waveform and parametric data appropriate to that interval to the SAS. The user will then interact with that data, making additions to the event and/or arrivals list by manipulating and processing the waveform and parametric data. These trial solutions and changes will be kept locally until the analyst has completed his processing of this time period. Then the analyst will send his results to the Parameter Database for appropriate updating of the parameter data. The SAS can also be used to define new functions for processing the data and testing out these functions, and provide for the analysts to keep private programs and data for their own use.

The Lincoln Laboratory is developing a prototype of the Seismic Analysis Station using a high capacity, high resolution vector graphics display, two high capacity alpha-numeric displays, a keyboard, data tablet, joy-stick, knob box, minicomputer, disk, tape, and other peripherals to

test out and refine the design concepts of the SAS. This prototype will be used to test procedures of event processing and interactive waveform processing in the environment of a Seismic Data Center.

Figure 8 shows a possible physical configuration of a work station for seismic analysts and researchers to interact with parameter and waveform data. The actual physical layout is subject to development from experience gained with the prototype.

The analyst or researcher will need several kinds of data displays. At least four data displays have been identified as potentially useful. These displays are: the events list whose entries characterize the events already identified, the arrivals list whose entries characterize the seismic wave arrivals at specific stations, waveform data, and geographic data such as maps. These displays will be shown on up to four screens as illustrated in Figure 9. The actual configuration including the number of screens is subject to the experience gained developing the prototype.

VII.B. Database Subsystem

The Database Subsystem consists of two major data management subsystems, the parameter database and the waveform database. The Parameter Database is responsible for all alphanumeric data as well as the reference event database. The Waveform Database is the other major data management subsystem. The Waveform Database stores all the received digital waveform data. A subset of the waveform data, which is in active use, is stored on

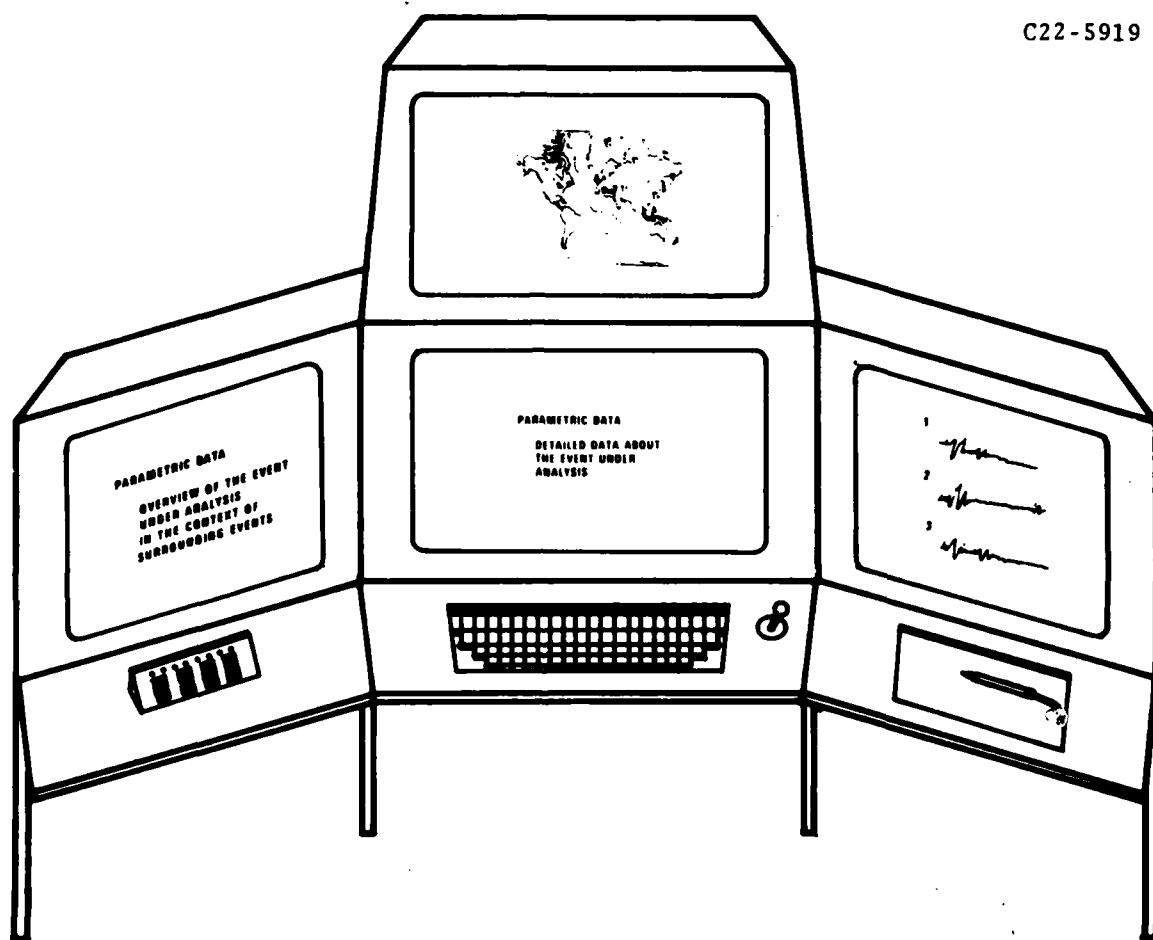


Fig. 8. Potential physical layout of a seismic analysis station.

C22-S920

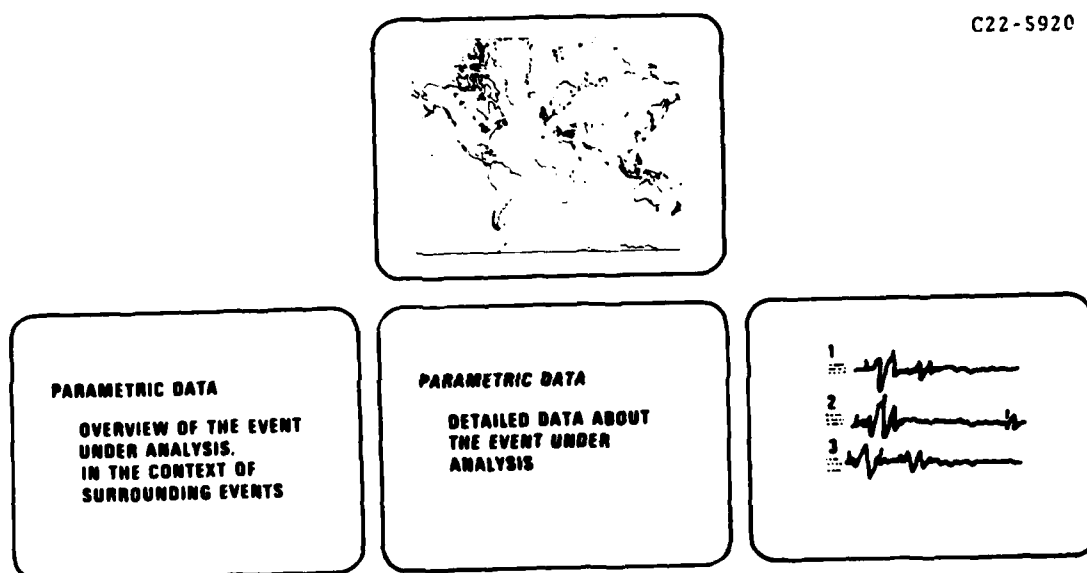


Fig. 9. Example of data displays.

disks for immediate retrieval. All the waveform data is stored in time segment magnetic tapes for retrieval as requested by users. The philosophy of operation and an overview of the design follow.

The Parameter Database Subsystem receives all the parameter data in the system, organizes it and makes it accessible for retrieval requests within the system.

The two principal classes of parameter data are the arrivals list and the events list. The arrivals list includes all of the arrival data reports received from the IESD and elsewhere, as well as all of the arrivals which are generated by seismic analysts within the system using the telemetered digital waveform data. The Automatic Processing Subsystem will provide a preliminary association of arrivals into events. An analyst will use all the arrivals, the preliminary event location data, and waveform data to determine the final characterization of the event which is then stored in the events list of the Parameter Database.

The events list will contain a record of each event found by the system. It may contain events imported from other systems such as the National Earthquake Information Service of the US Geological Survey. Each record will contain the parameterization of the event, such as the location, time, magnitude and other parameters. Each event entry will allow the retrieval of the associated arrivals which went into the characterization of the event.

The Parameter Database provides for the storage of these two time-ordered lists of parameter data, the arrivals and the events, and responds to requests for retrieval from these two lists by either time interval, list of event numbers, or list of arrival numbers. Mechanisms will be available for sifting either of these lists according to various criteria. The Parameter Database provides for storage of a list of data which has been received but not processed by the analysts. This list will include all the segments of data in which a seismic wave arrival was algorithmically detected and all the long period data. The Parameter Database also stores the master copy of all the various seismic tables used by analysis programs in the system. It is assumed that all needed tables will be stored to use in each computer for normal use. The master copy of each table is kept in the Parameter Database for retrieval by any computer needing to replace or update a local copy.

The Parameter Database can be thought of as residing primarily on-line, but for purposes of error recovery and integrity, a tape copy of all the data in the Parameter Database will be made on a routine basis.

The Parameter Database also has the capability to maintain a database of reference events. These reference events are collections of all the event, arrival and waveform data in the system for a particular set of events. The waveforms in a reference event are not continuous, but rather represent the segments of data containing the arrivals from specific events. Reference event data will normally be stored on the tape, but a

useful subset of these events will be kept on disk for immediate reference by analysts in their waveform processing operations. These reference events will primarily be those that are used as representative waveforms from some particular region for the purpose of arrival processing.

The parameter database manages lists of data entries corresponding to arrivals, events and reference events. The waveform database manages a totally different kind of data organization. The waveform database deals with a relatively small number of conceptually infinite length time series of digital waveform data samples.

The function of the Waveform Database is to store and retrieve the continuous waveform data, making it available for retrieval by time and station identification, ensuring that any segment of data is retrievable within a short period of time, normally in the range 5-15 minutes.

The continuous Waveform Database can be thought of as a virtual memory with all data being accessible by time and station. Reference to waveforms by event number will require the retrieval of a list of event numbers from the Parameter Database according to whatever selection criterion is desired. This list of events will contain the list of arrivals which went to make up the events. The arrival records will record the times of the waveform data (if any) which contains the arrivals. The program can then request this list of waveform data from the Waveform Database. The user will not need to be concerned with these details since the programs which

provide the user interface will handle all of the intermediate steps in fulfilling a request for data retrieval.

The primary storage of waveforms in the continuous Waveform Database is on magnetic tape. As it is received, the data will also be stored on disks for rapid and convenient accessing. The data is captured on tape at the earliest convenient moment, when the telemetered data will have all been received. The tape covering any given period will, therefore, normally be written within a few hours after the communicated data is received for that period. Late arriving data will be added to the tape at scheduled times. The data which has been saved on tape will still be kept on disk until the disk space is required for other data, at which time it will be overwritten. A user request for a specific piece of waveform data will be answered from disk whenever possible. If the data is not available on disk, an automatic request for tape retrieval will be issued, and the user so informed. The SDC design will have enough disk space to accommodate all the data that is in active use, otherwise there will be excessive demand on the use of the tapes to restore the data.

The disk and tape units which actually store the data for the waveform database will be connected to pairs of minicomputers to prevent a single failure of a computer or controller from destroying or preventing access to the data being managed by that computer. A representative configuration of the database module is given in Figure 10. This figure shows both disk and tapes on the same module. Normally in a large database configuration, some

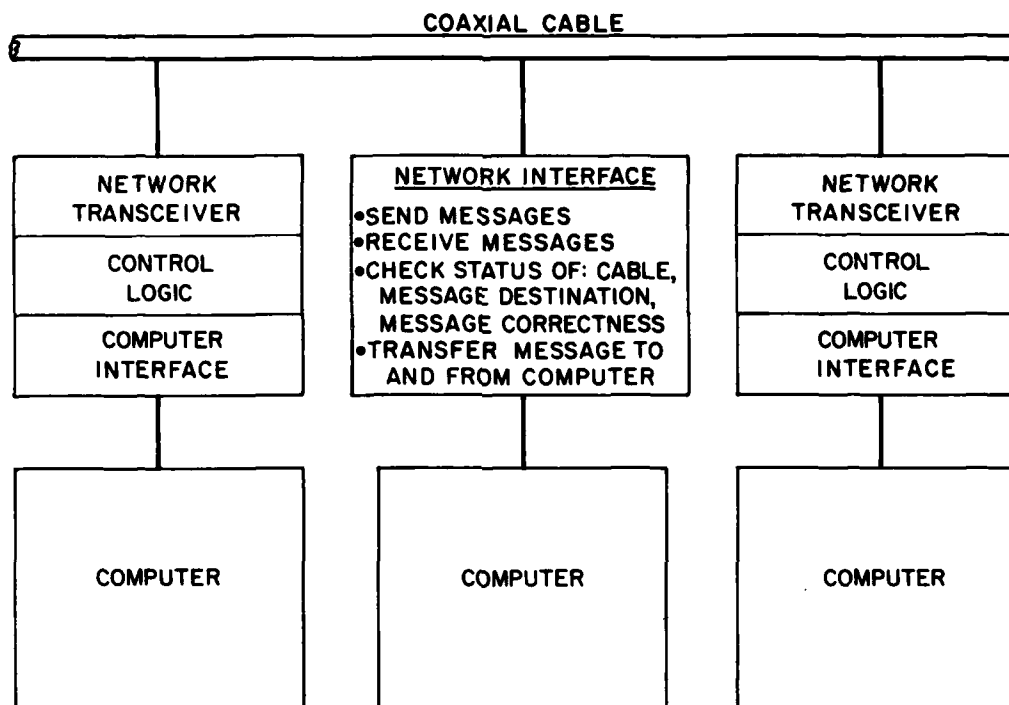


Fig. 10. Generic reliable data base module.

modules will have only disks and other modules will have tapes. All these computer modules will be connected to the local computer network for communication and control.

All requests for service from the waveform database will be addressed to the index manager of the database. The index manager is a routine which keeps track of the location of all the data in the system including whether or not the data is currently on disk or must be retrieved from tape. For each request, the index manager will pass along the request to the computer module which has access to the data either via disk or tape. The appropriate computer module will then transfer the data to computer originally making the request.

The Lincoln Laboratory is currently putting together a prototype of the redundant disk and tape module which will serve to develop the necessary programs and to demonstrate the operation of the waveform database. This is the minimum configuration for the waveform database. The expansion capabilities are realized by adding more modules, generally specialized to either disk or tape, up to the limits of the local computer network.

The parameter database is also being developed. Initially, the prototype parameter database will run on the research support subsystem computer and the SAS.

VII.C. Research Support Subsystem

The Research Support Subsystem will provide a variety of support services to seismic research users both local and remote. It will provide a multiple user UNIX environment for data display and analysis, system software development, a data and program storage facility for system users, and a means for entering certain types of user requests into the system. It will support both alphanumeric and graphics terminals for interactive data analysis and program development. It will provide dial-up access for authorized remote users to access data and program resources.

In the initial prototype, Lincoln Laboratory has acquired a single VAX-11/780 (tm Digital Equipment Corp.) computer to operate the Research Support Subsystem, the System Control and Services Subsystem and the Automatic Processing Subsystem. This computer will use the UNIX multiple-user operating system to support a variety of peripheral devices, including disk storage, tape I/O, and various alphanumeric and display terminals. The subsystem is expandable, if the system computational load requires, by the addition of further processing units, which may be for specialized tasks. For example, when a sophisticated signal processing algorithm is adopted, an array processor will probably be added to the automatic processing subsystem to support this algorithm and ensure that all the data can be processed in real time.

VII.D. System Control and Services Subsystem

The System Control and Services Subsystem will provide a variety of support services to the system as a whole. It will have two main functions. The first will be to monitor overall SDC status, and provide facilities to the system operator so that he can respond to operational problems. The second will be to provide a multiple user UNIX environment for system software development, test and analysis, a data and program storage facility for system users, and a means for entering certain types of user requests into the system.

In its initial configuration, the System Control and Services Subsystem will operate on the VAX-11/780 shared with the Research Subsystem and the Automatic Processing Subsystem using the UNIX multiple-user operating system. The subsystem is expandable, if the system computational load requires, by the addition of computers up to the limit of the local computer network.

VII.E. Automatic Processing Subsystem

The Automatic Processing Subsystem will provide for all non-interactive computational services. This will primarily be to provide computational power for certain processing algorithms, which will probably include automatic waveform processing and automatic event association.

In its initial configuration, the Automatic Processing Subsystem will

operate on the VAX-11/780 shared with the Research Subsystem and the System Control and Services Subsystem using the UNIX multiple-user operating system. The subsystem is expandable, if the system computational load requires, by the addition of computers up to the limit of the local computer network. If a sophisticated detection algorithm is adopted, one of the computers likely to be added would be one supporting an array processor capable of very fast processing of matrix operations.

VII.F. Communications Interface Subsystem

The Communications Interface Subsystem provides the input/output capability between the SDC and the outside world. There are three main functions: input of telemetered digital waveform data, input of magnetic tape data from all sources and the input/output of low speed alphanumeric data from/to the WMO/GTS network. Each of these functions will be elaborated in subsequent discussions.

The Communications Interface Subsystem's most demanding job is to acquire all of the waveform data which arrives via on-line telemetry. The Subsystem will also have the capability to receive waveform data on magnetic tape.

The Communications Interface Subsystem is also responsible for receiving and transmitting parameter data over the low speed communication lines. These low speed communications lines will be used for reception of the IESD data and for the transmission of the IESD data reports and event bulletins

generated by the SDC.

The initial prototype being developed by Lincoln Laboratory includes a communications interface subsystem implemented in two parts. The first part, the low-speed interface to the WMO/GTS is via a leased line to the VAX 11/780 used to support the Research, System Control and Services and Automatic Processing Subsystems. The reception of telemetered waveform data is planned for next year with an interface supported by minicomputers. This interface design is not yet completed.

VII.G. Local Computer Network

The local computer network is the subsystem which ties all the other subsystems together. Its operation is crucial to the success of the multiple minicomputer architecture. The following is an overview of the operation of the local computer network.

The local computer network consists of a passive data communications medium which interconnects the interfaces for the individual computers in such a way that any computer can broadcast information to all other computers at the same time. Each interface has the responsibility for copying all of the messages addressed to the host to which it is connected. In the design used here, the communications medium is a coaxial cable, the computer interfaces broadcast and receive over this cable, and the control of the communications is effected by microprocessors in the interface unit for each computer. Figure 11 shows the organization and representative

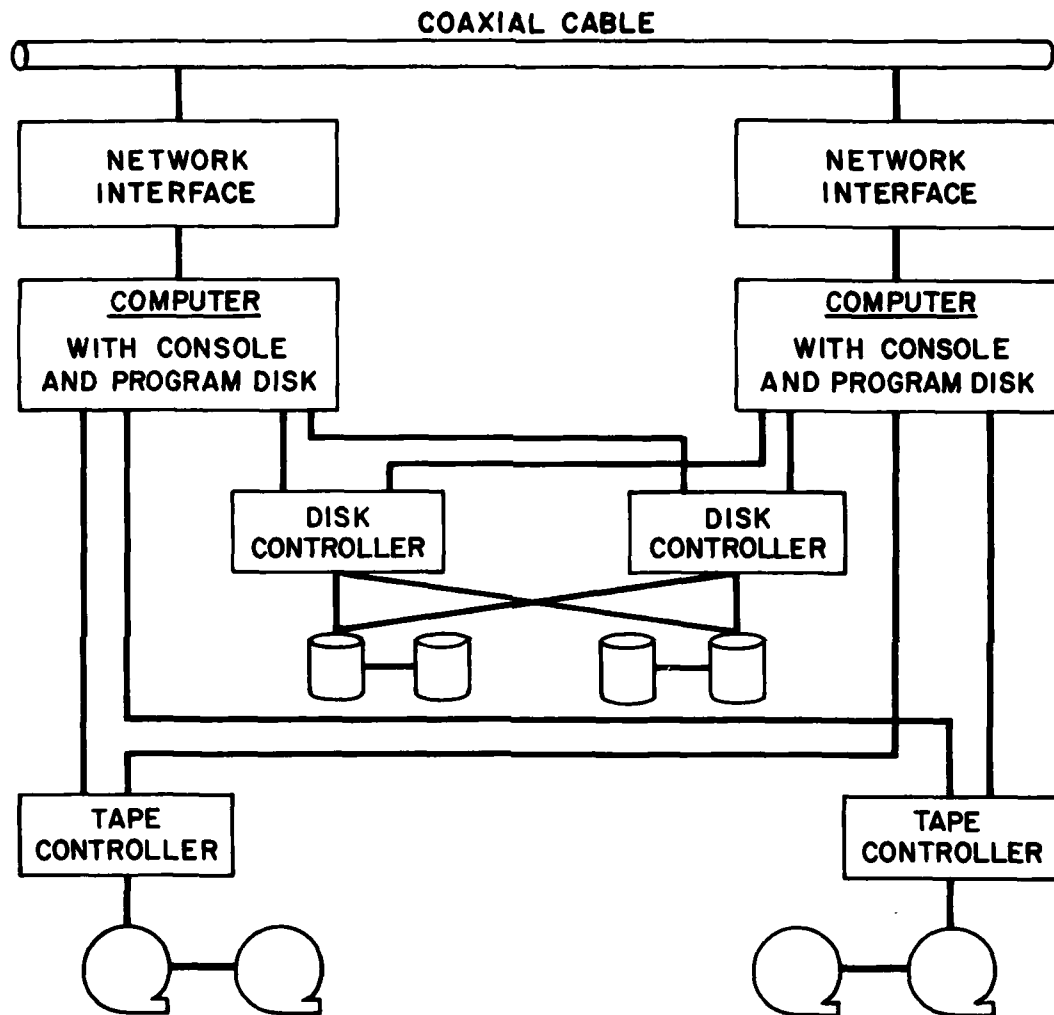


Fig. 11. Local computer network subsystem.

functions of the local computer network interfaces.

The local computer network interfaces is composed of three functional parts, the network transceiver, the control logic and the computer interface. The network transceiver is a straight-forward frequency shift keyed (FSK) transmitter and a receiver for the FSK signals. The control logic includes the parallel to serial conversion, the address recognition and status monitoring logic. The computer interface provides a parallel, direct memory access path for the messages to and from the host computer main memory. This part of the interface is the only part that has to be changed to accommodate a different computer on the local computer network. This design decision and the decision to implement much of the protocol logic in the control logic portion of the network interface greatly simplify adding of a new type of computer to the local network.

The operation of the system is based on a set of rules called protocols which allows the computers to carry on both coordination and command transmissions as well as block data transfers between computers on the network. There are a number of protocols that can be used on a broadcast network. Much analytic work has been done on these protocols. The "listen-while-talk" protocol chosen for this local computer network has been shown to be the most efficient, and it will allow the network to function smoothly and reliably with any load up to the capacity of the communication system. Additional levels of protocol to accomplish the reliable transfer of data streams and the exchange of command and status messages between

host computers are implemented by the network interface, thus relieving the host computers of some of the communications overhead.

The prototype being developed by Lincoln Laboratory includes a local computer network using CATV (cable access television) technology with vhf frequency shift keying as the communication technique and microprocessor based interfacing and protocol support for attaching the computers to the local computer network.

(16) DD60

(14) TN-1980-39

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER (18) <u>ESD-TR-80-97</u> ✓	2. GOVT ACCESSION NO. <u>AD A091621</u>	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) <u>Design and Development of a Seismic Data Center.</u>	5. TYPE OF REPORT & PERIOD COVERED (9) <u>Technical Note</u>	6. PERFORMING ORG. REPORT NUMBER <u>Technical Note 1980-39</u>
7. AUTHOR(s) (10) <u>Michael A. Chinnery and Alvin G. Gann</u>	8. CONTRACT OR GRANT NUMBER(s) (15) <u>F19628-80-C-0002</u> <u>WARPA Order-512</u>	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS <u>Program Element No. 61101E</u> <u>Project No. 0D60</u> <u>ARPA Order 512</u>
9. PERFORMING ORGANIZATION NAME AND ADDRESS <u>Lincoln Laboratory, M.I.T.</u> <u>P.O. Box 73</u> <u>Lexington, MA 02173</u> ✓	11. CONTROLLING OFFICE NAME AND ADDRESS <u>Defense Advanced Research Projects Agency</u> <u>1400 Wilson Boulevard</u> <u>Arlington, VA 22209</u>	12. REPORT DATE (11) <u>11 August 1980</u>
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) <u>Electronic Systems Division</u> <u>Hanscom AFB</u> <u>Bedford, MA 01731</u>	15. SECURITY CLASS. (of this report) (12) <u>48</u> <u>Unclassified</u>	15a. DECLASSIFICATION DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) <u>Approved for public release; distribution unlimited.</u>		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES <u>None</u>		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) <u>digital seismic data</u> <u>global seismic monitoring</u> <u>seismic data exchange</u>		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) <p>The U.S. has embarked on the design and development of a sophisticated data center for the analysis and management of seismic data. There are two broad motivations for this project. First, we need to develop a new generation system for the management and retrieval of digital seismic data, and to provide a modern data resource for the research community. Second, we need a facility that will be able to respond to any U.S. obligations that might be incurred under future agreements on international seismic data exchange and global seismic monitoring, such as those currently under discussion by the Committee on Disarmament. A computer architecture has been selected, consisting of a series of minicomputers connected by a high data rate local computer network. The principal components of the system are discussed, with emphasis on the data base management system, and the seismic analysis station, which provides an interactive man-machine interface.</p>		

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

207650

Am